

NONNEGATIVE DEFINITENESS OF THE ESTIMATED DISPERSION
MATRIX IN A MULTIVARIATE LINEAR MODEL

BY

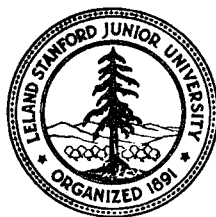
FRIEDRICH PUKELSHEIM and GEORGE P. H. STYAN

TECHNICAL REPORT NO. 32
MAY 1978

PREPARED UNDER CONTRACT N00014-75-C-0442
(NR-042-034)
OFFICE OF NAVAL RESEARCH

THEODORE W. ANDERSON, PROJECT DIRECTOR

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA



NONNEGATIVE DEFINITENESS OF THE ESTIMATED DISPERSION

MATRIX IN A MULTIVARIATE LINEAR MODEL

BY

FRIEDRICH PUKELSHEIM and GEORGE P. H. STYAN

Technical Report No. 32

May 1978

PREPARED UNDER CONTRACT N00014-75-C-0442

(NR-042-034)

OFFICE OF NAVAL RESEARCH

Theodore W. Anderson, Project Director

Reproduction in Whole or in Part is Permitted for
any Purpose of the United States Government

Approved for public release; distribution unlimited

DEPARTMENT OF STATISTICS

STANFORD UNIVERSITY

STANFORD, CALIFORNIA

Also issued as Technical Report No. 125 under National Science Foundation
Grant MPS 75-09450 - Department of Statistics, Stanford University.

Nonnegative Definiteness of the Estimated Dispersion
Matrix in a Multivariate Linear Model

by

Friedrich Pukelsheim
Universität Freiburg im Breisgau
and Stanford University

and

George P. H. Styan
McGill University

Summary.

Estimation is considered in a model where both the mean vector and the dispersion matrix have linear decompositions. It is shown that after an invariance reduction with respect to mean translation, MINQUE provides a nonnegative definite estimate of the dispersion matrix, when the decomposing matrices span a quadratic subspace of symmetric matrices. With normality, MINQUE is seen to equal the restricted maximum likelihood estimate and to be of uniformly minimum variance.

KEY WORDS: MINQUE. Noniterative solution of likelihood equations. Quadratic subspaces. REML. Special Jordan algebra. UMVU. Uniqueness of maximum likelihood estimate.

1. Introduction. Consider independent and identically distributed random \mathbb{R}^n -vectors \underline{y}_{α} , $\alpha = 1, \dots, N$, with common mean vector $\sum_{\pi=1}^p b_{\pi} \underline{x}_{\pi}$ and common dispersion matrix $\sum_{k=1}^k t_{k=k} \underline{v}_{k=k}$, where interest concentrates on estimating the vector $\underline{t} := (t_1, \dots, t_k)'$ of dispersion coefficients. Various procedures have been discussed in the literature. Among those are: (i) minimum norm unbiased quadratic invariant estimation (MINQUE, C.R. Rao [8, p. 302]), and, under normality, (ii) uniform minimum variance unbiased invariant estimation (UMVU, Seely [9]), and (iii) restricted (by invariance) maximum likelihood estimation (REML, Corbeil & Searle [2]). In this paper invariance is to be understood with respect to the group of all mean translations $\{\underline{y} \rightarrow \underline{y} + \sum b_{\pi} \underline{x}_{\pi} \mid (b_1, \dots, b_p)' \in \mathbb{R}^p\}$, a maximal invariant statistic being \underline{MY} where \underline{M} projects orthogonally onto the orthogonal complement of the space spanned by $\underline{x}_1, \dots, \underline{x}_p$; hence reduction by invariance yields the residual vectors \underline{MY}_{α} with mean $\underline{0}$ and dispersion matrix $\sum_{k=k}^k t_{k=k} \underline{MV}_{k=k} \underline{M}$.

Our main result may be roughly summarized as follows: If estimates according to each of the three procedures above exist, then they coincide, and the common estimate $\hat{\underline{t}}$ yields a nonnegative definite estimate $\sum_{k=k}^k \hat{t}_{k=k} \underline{MV}_{k=k} \underline{M}$ of the dispersion matrix in the invariance reduced model. This holds true for any finite sample size $N \geq v := \text{rank } \underline{M}$, in contrast to asymptotic results on consistency as $N \rightarrow \infty$, cf., Anderson [1].

In Section 2, the invariance reduced model is discussed in a normal setting, and Section 3 is concerned with the linear model situation.

The vital assumption is the condition of Seely [9] that

$\underline{\underline{M}}_1, \dots, \underline{\underline{M}}_k$ span a k -dimensional quadratic subspace \mathcal{B} of symmetric $n \times n$ matrices. The subspace \mathcal{B} is quadratic if and only if $\underline{\underline{A}}^2 \in \mathcal{B}$ whenever $\underline{\underline{A}} \in \mathcal{B}$, i.e., \mathcal{B} is closed under the multiplication

$\underline{\underline{A}} \circ \underline{\underline{B}} := \frac{1}{2}(\underline{\underline{A}}\underline{\underline{B}} + \underline{\underline{B}}\underline{\underline{A}})$. Jensen [4] points out that the latter property makes \mathcal{B} into a k -dimensional special Jordan algebra, and we shall adopt this more informative terminology. For a discussion with no initial invariance reduction, see Gnot, Klonecki & Zmyślony [3].

2. The Normal Model. We will use the isomorphism vec that maps a matrix into a vector by ordering its entries lexicographically, see Pukelsheim [7].

THEOREM 1. Consider independent and identically normally distributed random \mathbb{R}^v -vectors $\underline{\underline{Z}}_\alpha$ with common mean $\underline{\underline{0}}$ and common dispersion matrix $\sum_{k=1}^N \underline{\underline{W}}_k$, where $N \geq v$. Assume that the k decomposing matrices $\underline{\underline{W}}_k$ span a k -dimensional special Jordan algebra \mathcal{B} . Define $G \subset \mathbb{R}^k$ to be the region of those values $\underline{\underline{t}}$ of the dispersion parameter such that $\sum_{k=1}^N \underline{\underline{t}}_k \underline{\underline{W}}_k$ is positive definite, and assume $G \neq \emptyset$. Then:

(a) The maximum likelihood estimator for $\underline{\underline{t}} \in G$ is almost surely equal to the uniform minimum variance unbiased estimator

$$\hat{\underline{\underline{t}}} := (\underline{\underline{D}}' \underline{\underline{D}})^{-1} \underline{\underline{D}}' \cdot \text{vec} \underline{\underline{S}}, \text{ where } \underline{\underline{D}} := [\text{vec} \underline{\underline{W}}_1 : \dots : \text{vec} \underline{\underline{W}}_k], \text{ and}$$

$$\underline{\underline{S}} := \sum_{\alpha} \underline{\underline{Z}}_\alpha \underline{\underline{Z}}_\alpha' / N.$$

(b) The estimated dispersion matrix $\hat{\underline{\underline{W}}} := \sum_{k=1}^N \hat{\underline{\underline{t}}}_k \underline{\underline{W}}_k$ is nonnegative definite; in fact, if the sample dispersion matrix $\underline{\underline{S}}$ is positive definite, so is $\hat{\underline{\underline{W}}}$.

Proof. (a) Since G is open and connected it is a region, and its boundary ∂G consists of those $\underline{t} \in \mathbb{R}^k$ such that $\sum_{k=1}^k t_k \underline{W}_k$ is nonnegative definite and singular. The sample dispersion matrix \underline{S} is almost surely positive definite. If \underline{t} tends to ∂G , or $\|\underline{t}\|$ tends to ∞ , the likelihood function L tends to zero [1, p. 5]. Since L is positive in G there exists a maximum in G , and no maximum lies on the boundary ∂G . Hence the maximum likelihood estimate is a solution of the likelihood equations

$$(1) \quad \underline{D}' \underline{F}^{-1} \underline{D} \hat{\underline{t}} = \underline{D}' \underline{F}^{-1} \text{vec} \underline{S},$$

where the matrix of fourth moments

$$(2) \quad \underline{F} = \underline{F}(\hat{\underline{t}}) := (\sum_{k=1}^k \hat{t}_k \underline{W}_k) \otimes (\sum_{k=1}^k \hat{t}_k \underline{W}_k).$$

If \underline{F} in (1) is put equal to $\underline{F}(\underline{t}_0)$ for some given $\underline{t}_0 \in G$, then (1) is a set of weighted normal equations, cf. [7, p. 628], and hence yields a minimum variance unbiased estimator for the vector parameter \underline{t} . Since the matrices \underline{W}_k span a special Jordan algebra, there exists an almost surely unique *uniform* minimum variance unbiased invariant estimator which does not depend on the choice of $\underline{t}_0 \in G$. Thus

$$(3) \quad \hat{\underline{t}} = (\underline{D}' \underline{D})^{-1} \underline{D}' \text{vec} \underline{S},$$

since $G \neq \emptyset$ implies the existence of a nonsingular matrix $\underline{B} \in \mathcal{B}$, and so $\underline{B}^{-1} \in \mathcal{B}$ and $\underline{I}_v = \underline{B} \circ \underline{B}^{-1} \in \mathcal{B}$; the matrix \underline{F} in (1) may, therefore, be set equal to $\underline{I}_{v^2} = \underline{I}_v \otimes \underline{I}_v$.

(b) As a linear operator on the space of symmetric matrices, \hat{t} is surjective and hence open, and so if for some positive definite matrix \underline{S}_0 the value $\hat{t}(\underline{S}_0) \notin G$, the same is true for an open neighbourhood of \underline{S}_0 , i.e., for a set of positive Lebesgue measure. This contradicts part (a) that \hat{t} maps into G almost surely. For a singular sample dispersion matrix \underline{S} , consider the limit $\underline{S} + \epsilon \underline{I}_v$ as ϵ tends to zero. Q.E.D.

Part (a) may also be obtained from a reparametrization by $\underline{\theta} = \underline{\theta}(\underline{t})$, where the bijection $\underline{\theta}$ from G onto G solves $\Sigma_{\underline{\theta}}(\underline{t}) \underline{W} = (\Sigma_{\underline{t}} \underline{W})^{-1}$, as introduced by Seely [9, p. 715]. In this case one obtains an exponential family in the vector parameter $\underline{\theta}$ and standard theory applies, cf. Anderson [1]. A theorem proved by Mäkeläinen, Schmidt & Styan [6] may be used to obtain uniqueness of the solution to the likelihood equations (1).

3. The Multivariate Linear Model. We now return to the linear, but not necessarily normal, model discussed in Section 1.

THEOREM 2. Consider independent and identically distributed random \mathbb{R}^n -vectors \underline{Y}_{α} , $\alpha = 1, \dots, N$, with common mean vector $\Sigma \underline{b}_{\pi} \underline{x}_{\pi}$ and common dispersion matrix $\Sigma \underline{t}_{\underline{K}} \underline{V}_{\underline{K}}$, where $N \geq v = \text{rank } \underline{M}$. Assume that the k matrices $\underline{M} \underline{V}_{\underline{K}} \underline{M}$ span a k -dimensional special Jordan algebra \mathcal{B} that contains \underline{M} . Let $\underline{D}_{\underline{M}} := [\text{vec}_{\underline{M}}^{\underline{M} \underline{V}_{\underline{1}} \underline{M}} : \dots : \text{vec}_{\underline{M}}^{\underline{M} \underline{V}_{\underline{k}} \underline{M}}]$. Then the MINQUE

$$(4) \quad \hat{\underline{t}} = (\underline{D}_{\underline{M}}' \underline{D}_{\underline{M}})^{-1} \underline{D}_{\underline{M}}' \cdot \text{vec} \underline{S}$$

for \underline{t} yields a nonnegative definite estimate $\hat{\Sigma}_{\underline{t}}^{\underline{M}\underline{V}\underline{M}}$ of the invariance reduced dispersion matrix, this estimate being of rank ν if

$\underline{S} := \Sigma_{\underline{M}\underline{Y}_{\alpha}} (\underline{M}\underline{Y}_{\alpha})' / N$ is of rank ν .

Proof. It is easily checked that $\hat{\underline{t}}$ is the MINQUE in the enlarged model $[\underline{Y}_1' \underline{M} : \dots : \underline{Y}_N' \underline{M}]' \sim (\underline{0}, \Sigma_{\underline{t}} \underline{I}_{\underline{N}} \otimes \underline{M}\underline{V}\underline{M})$. The rest will be proved by reference to Theorem 1. Choose an $n \times \nu$ full rank ν factor \underline{Q} of \underline{M} , i.e., $\underline{M} = \underline{Q}\underline{Q}'$ and $\underline{Q}'\underline{Q} = \underline{I}_{\nu}$; then $\underline{Q}'\underline{Y}$ is another maximal invariant statistic [5, p. 707]. For the sole reason the proof, add a normality assumption. Then Theorem 1 is applicable to $\underline{Z}_{\alpha} := \underline{Q}'\underline{Y}_{\alpha}$, and yields the same $\hat{\underline{t}}$ as in (4); and the results on $\hat{\Sigma}_{\underline{t}}^{\underline{Q}'\underline{Y}\underline{Q}}$ imply the assertions on $\hat{\Sigma}_{\underline{t}}^{\underline{M}\underline{V}\underline{M}}$. Q.E.D.

If a normality assumption is added to Theorem 2, then using Theorem 1, we obtain the following:

COROLLARY. If the common distribution of $\underline{Y}_1, \dots, \underline{Y}_N$ is normal, then $\hat{\underline{t}}$ is the UMVU and REML estimate of \underline{t} , as well as the MINQUE.

Examples may be found in Corbeil and Searle [2]. In each one of their four cases a special Jordan algebra is present: equality of MINQUE (i.e., ANOVA estimators) and REML is implied by the Corollary and need not be checked explicitly, nor need the likelihood equations be solved iteratively.

4. Acknowledgements. This paper was presented at the Instytut Matematyczny PAN, Wrocław, by the first author, who would like to thank the Polish Academy of Sciences for their kind invitation. The authors would

also like to thank T.W. Anderson, K. Conradsen, and H. Witting for very helpful discussions. The second author's research was begun while he was a Canada Council fellow at the University of Helsinki.

INSTITUT FÜR MATHEMATISCHE STOCHASTIK, ALBERT-LUDWIGS-UNIVERSITÄT,
D-7800 FREIBURG IM BREISGAU, FEDERAL REPUBLIC OF GERMANY, AND DEPARTMENT
OF STATISTICS, STANFORD UNIVERSITY, STANFORD, CALIFORNIA 94305, USA.

DEPARTMENT OF MATHEMATICS, MCGILL UNIVERSITY, MONTREAL, QUEBEC,
CANADA H3A 2K6.

REFERENCES

- [1] T.W. Anderson, *Estimation of covariance matrices which are linear combinations or whose inverses are linear combinations of given matrices*, pp. 1-24 in: R.C. Bose, I.M. Chakravarti, P.C. Mahalanobis, C.R. Rao & K.J.C. Smith (eds.), *Essays in Probability and Statistics*, Statistical Publishing Society, Calcutta, 1969, and The University of North Carolina Press, Chapel Hill, 1970.
- [2] R.R. Corbeil & S.R. Searle, *A comparison of variance component estimators*, *Biometrics* 32 (1976), 779-791.
- [3] S. Gnot, W. Klonecki & R. Zymślony, *Uniformly minimum variance unbiased estimation in Euclidean vector spaces*, *Bull. Acad. Polon. Sci. Sér. Sci. Math. Astronom. Phys.* 24 (1976), 281-286.
- [4] S.T. Jensen, *Covariance hypotheses which are linear in both the covariance and the inverse covariance*, Preprint 1975, No.1, Institute of Mathematical Statistics, University of Copenhagen, January 1975.
- [5] J. Kleffe, *A note on MINQUE for normal models*, *Math. Operationsforsch. Statist.* 7 (1976), 707-714.
- [6] T. Mäkeläinen, K. Schmidt & G.P.H. Styan, *On the uniqueness of the maximum likelihood estimator of a vector valued parameter*, Discussion Paper No. 140, Indian Statistical Institute, Delhi, 1976.
- [7] F. Pukelsheim, *Estimating variance components in linear models*, *J. Multivariate Anal.* 6 (1976), 626-629.
- [8] C.R. Rao, *Linear Statistical Inference and Its Applications*, Second Edition, John Wiley & Sons, New York, 1973.
- [9] J. Seely, *Quadratic subspaces and completeness*, *Ann. Math. Statist.* 42 (1971), 710-721.

TECHNICAL REPORTS

OFFICE OF NAVAL RESEARCH CONTRACT N00014-67-A-0112-0030 (NR-042-034)

1. "Confidence Limits for the Expected Value of an Arbitrary Bounded Random Variable with a Continuous Distribution Function," T. W. Anderson, October 1, 1969.
2. "Efficient Estimation of Regression Coefficients in Time Series," T. W. Anderson, October 1, 1970.
3. "Determining the Appropriate Sample Size for Confidence Limits for a Proportion," T. W. Anderson and H. Burstein, October 15, 1970.
4. "Some General Results on Time-Ordered Classification," D. V. Hinkley, July 30, 1971.
5. "Tests for Randomness of Directions against Equatorial and Bimodal Alternatives," T. W. Anderson and M. A. Stephens, August 30, 1971.
6. "Estimation of Covariance Matrices with Linear Structure and Moving Average Processes of Finite Order," T. W. Anderson, October 29, 1971.
7. "The Stationarity of an Estimated Autoregressive Process," T. W. Anderson, November 15, 1971.
8. "On the Inverse of Some Covariance Matrices of Toeplitz Type," Raul Pedro Mentz, July 12, 1972.
9. "An Asymptotic Expansion of the Distribution of "Studentized" Classification Statistics," T. W. Anderson, September 10, 1972.
10. "Asymptotic Evaluation of the Probabilities of Misclassification by Linear Discriminant Functions," T. W. Anderson, September 28, 1972.
11. "Population Mixing Models and Clustering Algorithms," Stanley L. Sclove, February 1, 1973.
12. "Asymptotic Properties and Computation of Maximum Likelihood Estimates in the Mixed Model of the Analysis of Variance," John James Miller, November 21, 1973.
13. "Maximum Likelihood Estimation in the Birth-and-Death Process," Niels Keiding, November 28, 1973.
14. "Random Orthogonal Set Functions and Stochastic Models for the Gravity Potential of the Earth," Steffen L. Lauritzen, December 27, 1973.
15. "Maximum Likelihood Estimation of Parameters of an Autoregressive Process with Moving Average Residuals and Other Covariance Matrices with Linear Structure," T. W. Anderson, December, 1973.
16. "Note on a Case-Study in Box-Jenkins Seasonal Forecasting of Time Series," Steffen L. Lauritzen, April, 1974.

TECHNICAL REPORTS (continued)

17. "General Exponential Models for Discrete Observations,"
Steffen L. Lauritzen, May, 1974.
18. "On the Interrelationships among Sufficiency, Total Sufficiency and
Some Related Concepts," Steffen L. Lauritzen, June, 1974.
19. "Statistical Inference for Multiply Truncated Power Series Distributions,"
T. Cacoullos, September 30, 1974.

Office of Naval Research Contract N00014-75-C-0442 (NR-042-034)

20. "Estimation by Maximum Likelihood in Autoregressive Moving Average Models
in the Time and Frequency Domains," T. W. Anderson, June 1975.
21. "Asymptotic Properties of Some Estimators in Moving Average Models,"
Raul Pedro Mentz, September 8, 1975.
22. "On a Spectral Estimate Obtained by an Autoregressive Model Fitting,"
Mitutaki Huzii, February 1976.
23. "Estimating Means when Some Observations are Classified by Linear
Discriminant Function," Chien-Pai Han, April 1976.
24. "Panels and Time Series Analysis: Markov Chains and Autoregressive
Processes," T. W. Anderson, July 1976.
25. "Repeated Measurements on Autoregressive Processes," T. W. Anderson,
September 1976.
26. "The Recurrence Classification of Risk and Storage Processes,"
J. Michael Harrison and Sidney I. Resnick, September 1976.
27. "The Generalized Variance of a Stationary Autoregressive Process,"
T. W. Anderson and Raul P. Mentz, October 1976.
28. "Estimation of the Parameters of Finite Location and Scale Mixtures,"
Javad Behboodian, October 1976.
29. "Identification of Parameters by the Distribution of a Maximum
Random Variable," T. W. Anderson and S.G. Ghurye, November 1976.
30. "Discrimination Between Stationary Gaussian Processes, Large Sample
Results," Will Gersch, January 1977.
31. "Principal Components in the Nonnormal Case: The Test for Sphericity,"
Christine M. Waternaux, October 1977.
32. "Nonnegative Definiteness of the Estimated Dispersion Matrix in a
Multivariate Linear Model," F. Pukelsheim and George P.H. Styan, May 1978.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 32	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) NONNEGATIVE DEFINITENESS OF THE ESTIMATED DISPERSION MATRIX IN A MULTIVARIATE LINEAR MODEL		5. TYPE OF REPORT & PERIOD COVERED Technical Report
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) FRIEDRICH PUKELSHEIM and GEORGE P. H. STYAN		8. CONTRACT OR GRANT NUMBER(s) N00014-75-C-0442
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics Stanford University Stanford, California		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS (NR-042-034)
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Statistics & Probability Program Code 436 Arlington, Virginia 22217		12. REPORT DATE April 1978
		13. NUMBER OF PAGES 8
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report) Approved for public release; distribution unlimited.		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES Issued also as Technical Report No. 125 under National Science Foundation Grant MPS 75-09450 - Department of Statistics, Stanford University.		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) MINQUE. Noniterative solution of likelihood equations. Quadratic subspaces. REML. Special Jordan algebra. UMVU. Uniqueness of maximum likelihood estimate.		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) Estimation is considered in a model where both the mean vector and the dispersion matrix have linear decompositions. It is shown that after an invariance reduction with respect to mean translation, MINQUE provides a nonnegative definite estimate of the dispersion matrix, when the decomposing matrices span a quadratic subspace of symmetric matrices. With normality, MINQUE is seen to equal the restricted maximum likelihood estimate and to be of uniformly minimum variance.		

UNCLASSIFIED